Overlapping minds and the Hedonic Calculus

<u>Abstract:</u> It may soon be possible for neurotechnology to connect two subjects' brains such that they share a single token mental state - a feeling of pleasure, displeasure, desire, frustration, or something else relevant to welfare. How will our moral frameworks have to adapt to accommodate this prospect? And if this sort of mental-state-sharing might already obtain in some or all human cases, as philosophers have argued, how should this possibility impact our moral thinking?

This question turns out to be extremely challenging, because reflecting on different candidate examples generates opposite intuitions: if two mostly-distinct people share a few mental states, it seems we should count the value of those states twice, but if two physically-distinct beings share their whole mental lives, it seems we should count the value of that life once. We suggest that these conflicting intuitions can be reconciled if the mental states that matter for welfare have a holistic character, in a way that is independently plausible. We close by drawing tentative conclusions about how we ought to think about the moral significance of shared mental states.

Section 1: Rejecting Exclusivity

Much of morality is about maximising welfare: increasing the sum total of happiness, interest-satisfaction, or other measures of wellbeing, while reducing the sum total of suffering, interest-frustration, or other measures of illbeing. Utilitarians think that all of morality is fundamentally about this, but even non-utilitarians can agree that morality is sometimes about making more people better off, or at least making fewer people worse off.

This endeavour seems to center on identifying certain sorts of mental states, and evaluating their importance based on factors like intensity and duration. And generally in our moral thinking, as elsewhere, we assume that each mental state belongs to exactly one person: a given pain can vary in intensity, duration, and other such ways, but cannot vary in how many people feel it. I feel my pain, you feel yours, and that is it. Call this assumption 'Exclusivity'. What if this assumption is wrong? What if, that is, there could be two or more subjects feeling the very same pain? A variety of arguments independently suggest that Exclusivity might cease to be true in the future, might already admit of exceptions, or might even fail to hold in everyday cases (section 2 reviews four examples).¹ None of these arguments are decisive, and metaphysical defences of Exclusivity can be and have been offered², but our interest here is in what would follow for morality if Exclusivity were relaxed or abandoned.³ The most basic question can be put crudely as: if two subjects feel the same token pain, is that twice as bad as only one subject feeling it, all else being equal? Or is it equally as bad as only one subject feeling it, all else the 'subject-counting' and 'state-counting' approaches⁴, and call this puzzle the 'Value-Counting' question.

Deciding the Value-Counting question might bear on the best way to allocate rights, adjudicate conflicts, or distribute scarce resources in situations with both connected and unconnected minds, and might also bear on which kinds of minds we should create or not create in the future. Advances in AI or neurotechnology might make it possible to engineer minds with structures and relationships very different from those typical of humans and nonhuman animals, and deciding in advance how to calculate welfare when mental states are shared may become crucial to making moral sense of such differently-structured minds.

¹ Key discussions include Unger 2004, Hirstein 2008, Zimmerman 2010, Langland-Hassan 2015, Blackmon 2016, Simon 2017, Cochrane 2021, Javier-Castallenos 2021, Kang 2022.

² For explicit discussion of this principle see esp. Dainton 2011, p.246, Kriegel 2017, p.131 ff, Roelofs 2019, p.60ff, Roelofs 2020.

³ Thus we are picking up on the thread left dangling by Sutton when, after making a metaphysical case for the failure of Exclusivity, she remarks that "how to figure a shared pain into a utilitarian calculus is an interesting but separate question that I will leave to the ethicists." (Sutton 2014, p.622).

⁴ These are generalisations of the views that Javier-Castellanos calls "Quantity of Pain" and "Number of Experiencers", 2021, p.5. Note that Sutton, right after setting aside the ethical question, remarks that her "first-pass response is to count by quantity of pains rather than number of beings sharing that pain", i.e. to endorse the state-counting approach (2014, p.622).

In the next section, we review four different contexts where exclusivity seems questionable and welfare-relevant mental states might be shared. Some of these cases prompt subject-counting intuitions, while others prompt state-counting intuitions. In sections 3 and 4 we ask what to make of these different intuitions and argue that the best way to reconcile them is to

endorse the state-counting approach but explain the subject-counting intuitions by appeal to the relatively holistic nature of hedonic experience and preference satisfaction. Because much or all of the value or disvalue of a mental state is not intrinsic to it, but consists in how it contributes to a total mental life, we can explain why small degrees of overlap do not, in practice, affect the separate valuation of the overlapping individuals.

The Value-Counting Question is our primary focus, but it potentially interacts with at least three other questions that will become more important as AI and neurotechnology advance. First, consider what we call the Hedonic Emergence question: Does hedonic valence in a complex experience entail hedonic valence in some parts of that experience? And vice versa, does hedonic valence in the parts entail hedonic valence in the whole? Or could valence be 'emergent', at least in the weak sense of belonging to composite experiences no component of which is itself valenced? The holistic position we defend in section 4 suggests a negative answer to the Hedonic Emergence question: the valence of a composite experience depends on its overall structure more than on the valenced or unvalenced status of its parts.

Second, consider what we call the Hedonic Alignment question: If one mind is composed of other minds, how hedonically aligned must these minds be? Might a collection of unhappy minds collectively constitute a happy mind (or vice versa), or can they collectively constitute only an unhappy mind? Some of the examples we discuss will turn crucially on whether there is any guarantee of alignment between parts and wholes, and how strong it might be. The holistic position we defend in section 4 suggests a nuanced answer to the Hedonic Alignment question: alignment is guaranteed only in proportion to the size of overlap relative to the larger mind, so a composite mind with many very small mental parts might diverge from them sharply in welfare, while one with only a few large mental parts will be more closely aligned.

Finally, the choice between subject-counting and state-counting approaches evokes the debate between person-affecting and impersonal views of welfare.⁵ Can a situation be morally better even if no identifiable person is better off in it? On impersonal views, the answer is yes: the situation might be better because it contains more happiness, equality, or some other sort of impersonal value. On a person-affecting view, the answer is no: the comparative value of situations must be tied to their comparative value for specific individuals. We will suggest that the subject-counting approach fits naturally with the person-affecting view, though this is not a strict entailment.

Section 2: Candidate Cases of State-Sharing

In this section we further explain and motivate the Value-Counting question by describing in more detail what kinds of mental connection might be possible and why these kinds of mental connection might matter. First, in subsections 2.1 and 2.2, we describe two general ways for minds to be connected: neural telepathy (where two mostly distinct subjects have partial overlap in their mental states, in virtue of an anatomical or technological connection) and hive minds (where one mind is composed by many others). Then, in subsections 2.3 and 2.4 we describe two

⁵ For discussion see, e.g., Arrhenius 2003, Temkin 2014, Bader 2022.

philosophical debates that turn in part on the possibility of connected minds: the Problem of the Many and the moral implications of constitutive panpsychism.

2.1 Neural Telepathy

First, welfare-relevant states might be shared if two brains are connected in ways that mimic the connections within a single brain - what we might call 'neural telepathy' or 'mindmelding' (Hirstein 2008, 2012, Danaher and Nyholm 2021, Lyreskog et al. 2023). This might result from technology known as a brain-to-brain interface, albeit likely from a version more advanced than anything currently on the horizon. This would involve using two connected brain-computer interfaces, linked to implants in two brains, allowing the brains to transmit information to one another in a way functionally equivalent to how nerve tracts transmit information from one brain area to another. If we suppose that the boundaries of the mind are primarily fixed by information-processing patterns, not by the anatomical boundaries of particular tissue types,⁶ this might result in a situation where the physical supervenience base of the one person's states (alternatively, it might be that connections themselves become part of the supervenience base for both subjects' states). This doesn't mean that the two people have thoroughly merged, or come to function as a single person.⁷ It just means that some subset of states now belongs to both of them.

There may be existing cases where atypical brain conditions already create something like neural telepathy. The split-brain condition is often described as creating a split between two distinct subjects of consciousness; if this is the case, then it seems likely that the extensive neural

⁶ This assumption is a very minimal form of the Extended Mind hypothesis, though it is compatible with denying that any currently-existing technology does or could extend the mind, if the functional criteria that would need to be met are sufficiently demanding. Cf. Clark and Chalmers 1998, Clark 2008, Vold 2015, Chalmers 2019.

⁷ Though this sort of technology might also be used to bring about that result, cf. Roelofs 2019, p.270 ff.

overlap between these two hemisphere-subjects (in the brainstem, thalamus, and other subcortical structures) might underlie some shared token states.⁸ However, the split-brain is in many ways an unusually puzzling case, because it seems at most times to involve only one subject - most notably, if there are two subjects here, they're not aware of each other, but seem to usually cooperate seamlessly, each taking themselves to be the unique subject present.⁹

A clearer candidate case would be craniopagus conjoined twins, who are distinct individuals joined at the head, potentially with linked or overlapping brains. The case of Canadian twins Krista and Tatiana Hogan has been discussed by philosophers as a possible case of state-sharing because of the 'thalamic bridge' that seems to allow sensory information to pass between them.¹⁰ And unlike in the split-brain, any state-sharing here would co-exist with clear awareness on the part of both subjects that they are two in number. An intriguing example is the report that taste information can be transferred, even while hedonic valence seems to differ:

The sensory exchange, [parents] believe, extends to the girls' taste buds: Krista likes ketchup, and Tatiana does not, something the family discovered when Tatiana tried to scrape the condiment off her own tongue, even when she was not eating it. (Dominus

2011)

If this incident involved neural telepathy in our sense, that would mean that a single token sensory experience (the taste of ketchup) was produced by Krista putting ketchup in her mouth, and then experienced by both her and Tatiana. Moreover, because of their different preferences, this taste experience was somehow part of an overall pleasant experience for Krista, but part of an overall unpleasant experience for Tatiana (a possibility that will be important in section 4).

⁸ See e.g. Gazzaniga and Sperry 1962, Schechter 2015, 2018, Pinto et al. 2017.

⁹ See esp. Schecter 2018, pp.156-180.

¹⁰ See esp. Dominus 2010, Hershenov 2013, Langland-Hassan 2015, Kriegel 2017, pp.132-133, Roelofs 2019, p.112n, Cochrane 2021, Kang 2022.

It is worth stressing that what we are calling 'neural telepathy', involving the sharing of token states, requires more than simply that information-transfer between brains causes both subjects to have distinct token states of the same type. If Krista eating ketchup caused two matching taste-states - one for her and one for Tatiana - that would not be neural telepathy in our sense. We might term the state-sharing version 'strong telepathy' and the version with type-identical duplicates 'weak telepathy': We take it that both strong and weak telepathy are possibilities in principle, and that in any given case it might be extremely hard to determine empirically whether a particular pair of subjects are an example of one or the other. It might be that no actual cases of strong neural telepathy exist at present or will exist in the future; nevertheless it is worth asking how that possibility, if and when it comes about, should affect moral theorising.

2.2 Hive Minds

Next, consider the possibility of 'hive minds': distributed intelligent systems, made up of individuals suitably networked or communicating with each other, where both the whole and the parts have minds, and some or all of the whole's mental states are ones 'inherited from' its various parts, and to that extent shared with them.

Some possible examples of hive minds exist at present. For an example concerning parts and wholes, consider octopuses. Octopuses are sometimes said to have 'nine brains', one central brain and another, smaller brain for each arm. They also exhibit some integrated behavior, where it appears that the organism is acting as a whole, and some fragmented behavior, where it appears that parts are acting independently of each other.¹¹ Assuming that consciousness exists in

¹¹ For discussion see, esp. Carls-Diamante 2017, 2022.

octopuses at all, is it possible that consciousness exists in each brain as well as in the octopus as a whole?

Alternatively, for an example concerning individuals and groups, consider colonies of eusocial insects like ants and termites. Members of these colonies can and do act independently of each other, but the colonies also exhibit a remarkable degree of functional unity. And it is still unknown whether these insects are conscious, individually or collectively.¹² Is it possible that consciousness exists in each insect as well as in the colony as a whole - and if so, do the experiences of each insect also form part of the colony's experiences, thus belonging simultaneously to both?

Other possible examples might arise in the future. Some might even involve humans as parts, as a result of the sort of neural telepathy described in the last subsection. There we focused on states being shared between two (or more) human brains, but what if the system constituted by these brains can be conscious as well? Then we would have a consciousness in each member of the group as well as in the group as a whole, in much the same way that we might currently have with some invertebrates (cf. Danaher and Nyholm 2021, Danaher and Petersen 2021). What might it be like to be a conscious group of humans, or to be a member of such a group?

Finally, a hive mind structure might appear in artificial conscious beings as well. If and when artificial intelligences become conscious, they might be linked in the same kind of way that invertebrate minds are, but to a much greater extent. In particular, a vast number and wide range of artificial minds might be linked with each other via the internet. In this case, it might even be possible for hive minds to be layered, with individual minds making up collective minds that,

¹² For discussion of insect consciousness, see Klein and Barron 2016; for discussion of the cognitive status of eusocial insect colonies, see Huebner 2014, and for discussion of their potential consciousness see Friedman and Søvik 2021.

themselves, make up even larger collective minds. If so, what would be the ethical status of these layered hive minds?

2.3 Puzzles of Multiplication

There might also be shared mental states between overlapping parts of one human being, such as their top half, their head, and their brain. Consider that if my severed head existed by itself (with appropriate life-support) it could be conscious; so could my brain, or my top half. So these entities are, considered intrinsically, capable of consciousness. And why would their current situation (connected to the rest of me) suppress or inhibit such consciousness? So it might seem to follow that they are likely to be conscious right now - but since they are apparently distinct from one another (and from me, the whole human being), this seems to imply that there are multiple conscious subjects right here and now. This has been called the problem of 'Too Many Minds'.¹³

There are other ways to generate multiplicity, such as through the gradedness of psychological continuity. One common reason for distinguishing a conscious subject from the physical system that realises it is that they seem to have different persistence conditions - perhaps the physical system persists through biological continuity, while the subject persists through psychological continuity. But how much psychological continuity is necessary? Various forms of psychological change, disruption, or transformation can occur, some that intuitively seem to destroy the previous person, some that intuitively seem not to, and some that seem borderline. We might reason our way to positing multiple co-located beings whose persistence requires different degrees of psychological continuity, e.g. a relatively internally consistent 'angry self' and 'calm self', as well as a less consistent persisting self perennially dismayed at how its angry

¹³ See e.g. Unger 1980, 2004, Burke 1994, Blackmon 2016, 2021

self and calm self keep undoing each others' actions.¹⁴ If we accept the coexistence of these beings, it would be hard to deny that the mental states belonging to each of the 'smaller' more consistent selves also belong to the 'larger' self that contains them both.

Different philosophers respond to puzzles like this in different ways; many try to avoid any overlapping subjects by denying various premises. For instance, to avoid saying that both I and my brain are conscious, someone might deny that the brain itself is conscious, even though it would be if it existed by itself. Alternatively, someone might say that I just am my brain, and that no larger whole containing that brain (like 'the whole human being') is really conscious. But both of these options are somewhat counter-intuitive, and in general philosophers inclined to think that some physical systems can have minds must grapple with the likelihood that such systems may often overlap with each other.

However, other philosophers, sometimes called 'manyists', embrace the implication that many different physical systems have minds, but maintain that this is not a problematic result.¹⁵ Because these many systems all overlap, they claim, they share the same token mental states, and are appropriately 'counted as one' collectively for most purposes. But evaluating the viability of manyism requires evaluating the metaphysical and moral implications of shared mental states, and critics of manyism have repeatedly suggested that accepting overlapping minds yields unacceptable revisions to everyday morality.¹⁶

¹⁵ See e.g. Lewis 1993, Sutton 2014, Roelofs 2022

¹⁴ See e.g. Olson 2003, Noonan 2010, Sebo 2015-a, 2015-b. For discussions specifically of how something like the value-counting question interacts with perdurantism about persistence, see Briggs and Nolan 2015, Johnston 2017, 2021, Javier-Castellanos 2021, cf. Schwitzgebel 2020.

¹⁶ See esp. Simon 2017, Johnston 2017, 2021, Javier-Castellanos 2021.

2.4 Constitutive Panpsychism

Finally, consider the world as envisaged by constitutive panpsychists. Here state-sharing might be incredibly widespread, indeed ubiquitous. Constitutive panpsychism maintains that the fundamental physical constituents of the universe have some very simple sort of consciousness, and that our complex consciousness is combined out of this somehow.¹⁷ On some accounts of how this combination works, it involves token states being shared: every token experience of yours is composed of simpler states, themselves composed of simpler states, down to the simplest phenomenal components, each of which is shared between you and some microscopic part of your brain. If constitutive panpsychism is true, there is a sense in which we may all be hive minds. Indeed, the supposed impossibility of state-sharing has been raised as a potential objection to constitutive panpsychism, part of the "combination problem."¹⁸

Our interest here, though, is not in the combination problem *per se*. Rather, it is in how debates about mental combination bear on the *moral* implications, if any, of panpsychism. If my pleasures and displeasures are constituted by states of my smaller parts, what does that tell me about the pleasures and displeasures of those smaller parts? This seems to turn on what in section 1 we called the Hedonic Emergence and Hedonic Alignment questions. And if my parts do have valenced experience, what does that mean morally? This raises the question on which we are focusing here: the Value-Counting question. So determining the moral implications of panpsychism partly requires answering this question.¹⁹ To put it another way, if constitutive panpsychism is true, then we may have to apply whatever conclusion we reach about the moral status of hive minds to ourselves as well.

¹⁷ See e.g. Seager 1995, Chalmers 2015.

¹⁸ See e.g. Roelofs 2016, pp.3205-3209, 2019, 2020, pp.61-69, Miller 2017, p.12.

¹⁹ For other work in this area see Mathews 2003, Buchanan and Roelofs 2019, Vetlesen 2019.

Section 3: The Tension among Intuitions

We think intuitive reactions to these different cases are likely to point in opposite directions.²⁰ In considering neural telepathy, and perhaps also when considering hive minds, subject-counting seems intuitively the right approach. At least this is the case when the proportion of shared states is small. Suppose, for example, that some neurotechnological intervention brought it about that you and your friend shared the tactile states associated with your left hand, but nothing else. Suppose, that is, not just that stimulation of that hand produces matching states in both of you, but that it produces a single experience - a single complex of neural activity - that is undergone by both of you, because it occurs in a neural region that is integrated equally into both of your brains. Now suppose someone hits that hand with a hammer. This causes you intense pain (which is bad), and moreover it causes your friend intense pain (which is also bad). Intuitively, this situation is similarly bad, hedonically speaking, as a situation where two discrete people both have their hands hit with hammers. The fact that, by hypothesis, the imagined situation involves a single shared pain rather than two distinct pains seems beside the point: what matters is that two people are both in severe pain. To count the pain only once would seem unfair to you and your friend: it might mean, for instance, that scarce pain relief is preferentially allocated to two unconnected people each in moderate pain, rather than to the two of you, both in severe pain.

But in discussions of the puzzle of multiplication or constitutive panpsychism, subject-counting shared pains threatens absurdity. For example, if one person is physically larger than another, there will be many more distinct subsets of their atomic parts, and thus potentially more conscious subjects in their vicinity: if each subject has equal moral status to the others, then

²⁰ This tension is a generalisation of the argument of Javier-Castellanos (2021, p.5 ff), who argues against views of personal identity that imply a multiplicity of subjects in one place, precisely because (he claims) they require a state-counting approach to avoid bizarre moral implications; the state-counting approach, Javier-Castellanos then points out, appears to generate the wrong moral judgements in cases involving craniopagus conjoined twins.

causing any sort of suffering to the larger person (i.e. the larger 'cloud' of subjects) will be perhaps millions of times more serious than causing similarly intense suffering to the smaller person (i.e. the smaller 'cloud' of subjects). Indeed, critics of manyism have raised precisely this worry; Simon (2017) calls this 'the hard problem of the many', and offers it as a *reductio* of manyism and by extension of materialism itself.²¹ Roelofs (2022) responds that these results need not follow, if mental states can be shared (cf. Sutton 2014). But that response relies on the intuitive appeal of state-counting the value of shared states: if I hit my hand with a hammer, for instance, and cause myself intense pain (which is bad), it seems misguided to say that the situation is even worse because my head is also in intense pain (and also my top half, and all-of-me-except-one-atom, and so on): it's just the same pain, appearing again in a different guise. Hence defenders of manyism can insist that the many conscious subjects can matter *equally* without having to matter *in addition* to each other - *if* they can maintain the state-counting, rather than subject-counting, approach.

We might resolve this tension in relatively conservative ways, by denying that welfare-relevant states are shareable, either in general or in particular cases. For instance, if we simply affirm Exclusivity (perhaps viewing the discussion so far as offering a *reductio* of state-sharing) then the puzzle does not arise. This might be especially plausible if we accepted substance dualism, and if we stipulated that souls are both perfectly indivisible (sidestepping the puzzles of multiplication) and incapable of direct interaction or contact (ruling out strong neural telepathy). Indeed, Simon (2017) offers the moral objection to manyism specifically as part of an argument in favor of substance dualism (see also Unger 2004, Zimmerman 2010).

Alternatively, if we deny that neural telepathy and hive minds can produce shared mental states, then we can accommodate our intuitions about these cases without accepting

²¹ See also e.g. Zimmerman 2010, Johnston 2017, 2021.

subject-counting. This might be especially plausible if we accepted the exclusion postulate of Integrated Information Theory (see e.g. Oizumi et al. 2014, Mørch 2019, Albantakis et al. 2022), according to which as soon as two minds become intimately enough connected to share an experience, they immediately cease to exist as two minds and instead simply exist as one.

Finally, if we deny that humans generally overlap with huge numbers of other subjects, then we address the puzzle of multiplication without accepting state-counting. This might be especially plausible if we thought that we could solve these puzzles in a non-manyist way, by showing that there really is only one entity present that qualifies as a subject.

We will not argue against these relatively conservative solutions: Instead, we will explore a more liberal solution, which embraces the possibility of mental overlap across a range of cases and tries to reconcile the disparate intuitions this generates. We have several reasons for developing this more liberal approach. One is that we personally feel skeptical that the benefits of these conservative approaches will outweigh their various costs. Another is that we are interested in knowing what *would* follow ethically if state sharing were metaphysically possible. And, another is that we think that the liberal solution not only resolves the tension discussed here but also clarifies its relation to a range of other metaphysical and ethical issues, as we will see in section 5.

It might also be that everyday morality reflects a commitment to Exclusivity and/or substance dualism. The charge that utilitarianism does not respect 'the separateness of persons' (Gauthier 1962, p.126, Rawls 1971, p.27, Nozick 1974, p.33; cf. Hinton 2009) might seem to imply that persons are necessarily, and sharply, separate, and cannot share any of their mental states with one another.²² If so, then we might think that the impossibility of state-sharing is one

²² Compare Kriegel 2017, who links "Phenomenal Inviolability" (p.131; equivalent to our principle of Exclusivity) with the "normative inviolability" of persons (p.133).

of the key facts about the world that our ethical theories should reflect. But we could equally flip this point around by concluding that if two subjects *could* share an experience, or other mental state, then our ethical thinking would have to be updated or revised to reflect that possibility.²³ Someone might insist that we can rule out such a possibility precisely because it would go against such a deep ethical truth. But it seems unwise to put limits on what is metaphysically possible based on ethical convictions that might, themselves, depend on false metaphysical assumptions.

Section 4: Resolving the Tension Through Holism About Welfare-Relevant States

The ideal view, for someone who thinks welfare-relevant mental states can be shared, would be one which, on the one hand, supports subject-counting in cases of only *slight* mental overlap, since these are the kinds of cases in which our intuitions most support subject-counting, but which, on the other hand, supports state-counting in cases of near *total* overlap, since these are the cases in which our intuitions support state-counting. We think that a view of this sort is available, through affirming state-counting while stressing the holistic character of welfare-relevant mental states. That is, when two mostly distinct subjects share just a few states, we can hold that the value that is strictly intrinsic to those states should still be state-counted, while also holding that most or all of their value is not intrinsic to them, but rather consists in their contributions to each subject's larger field of unshared mental states.

Recall the ketchup case noted above, where a potentially shared taste experience was pleasant to one of the Hogan twins but unpleasant to the other. Here there is no temptation to say

²³ This move is in some ways parallel to Parfit's argument (1984, pp. 329-346) that since persons are not actually separate in any deep or fundamental sense (assuming the falsity of substance dualism), utilitarianism is more defensible than it might otherwise be.

that the ketchup taste itself is intrinsically good or bad: value comes from the unshared valenced states into which the taste factors. On the holistic proposal, this provides a model for shared states in general. For example, in the case where a hammer hitting your hand causes pain to both you and your technologically-linked friend, we can say that the experience of the hammer hitting your hand is not itself the bearer of disvalue, but is rather the basis for (intense) disvalue in your total experience.²⁴

A strong version of this claim might go as far as to deny that *any* particular component of your experience has valence, considered in and of itself. On this radically holistic view, what we have been describing as a pain is not, in and of itself, negatively valenced at all, but is rather a type of sensation that strongly tends to make any total experience that contains it negatively valenced. On this view, when we speak of particular experiences as 'pleasant' or 'unpleasant', this is only as a shorthand for 'typically makes the total experience containing it more pleasant/unpleasant'. So although your suffering and your friend's suffering are both based partly on the shared experience, they are each unshared, and we should count them separately when estimating the overall badness of the situation.

A weaker version of the holistic claim would allow that what we have been describing as a pain is *somewhat* negatively valenced in and of itself, but would say that this is not where most of its disvalue lies. Rather, most of its disvalue lies in its contribution to your total experience. For instance, we might point to the causal effects that this pain has on your other, unshared, mental states, such as how it frustrates your desire not to experience sensations like that and fills your mind with memories of past pains, anticipations of future pains, or associated emotions

²⁴ Though we cannot here explore all its consequences, a parallel holistic claim about the diachronic composition of lifetime welfare - that, e.g., the same moments might add up to a better life if they occurred in one order than in another - is defended by Slote (1982) and Velleman (1991); Briggs and Nolan discuss how this might complicate various approaches to the puzzles of diachronic multiplication (2015, pp.404).

such as fear and anxiety about how the pain might interfere with your projects or relationships. And we might say that the negative valence of these further, unshared states outweighs the negative valence of the pain that led to them.

However exactly we spell out the holistic claim, it implies that a welfare evaluation of the situation will be dominated by what is not shared between your friend and you, and thus will vindicate the intuition that causing an intense shared pain is as bad as causing two unconnected subjects intense pains, *even if* we assume state-counting. That is, subject-counting intuitions can make sense in practice even if state-counting is true in principle. This includes cases involving neural telepathy, since in these cases, what is unshared matters much more than what is shared, and so it makes sense to treat these cases as involving separate subjects for many purposes. But conversely, in the cases generated by manyism, everything is shared, and so it makes sense to treat these uncertained by manyism.

The vindication of subject-counting intuitions may be more or less complete, depending on exactly how strong the holistic claim is made. Weaker sorts of holism might still imply that the shared-pain situation is slightly less bad than the situation with two unshared pains: this may be a theoretical cost, though it is unclear how fine-grained our moral judgements really are about this sort of case. Given the epistemic and theoretical difficulties with quantifying and comparing phenomenal characters, we might be happy with views that give the right rules of thumb.²⁵

²⁵ The split-brain case stands out as intermediate between only slight overlap (supporting the subject-counting intuitions) and complete overlap (supporting the state-counting intuitions). There are important connections not just among specific sensations but major psychological structures, such as agency and self-consciousness. If there are in fact two distinct subjects here, it is very unclear what a holistic view of the value of states would imply, since much of the overall shape of their total states seems to be either shared or at least kept synchroniz*f* ed somehow. A strong holist might still say that these two subjects should be treated as separate and additional subjects in our moral calculations, while weaker versions are much harder to apply. We suggest that the moral upshot of holism here is deeply confusing and unclear, and that this is exactly the right result to get: the split-brain case does indeed generate deeply confusing and unclear reactions, insofar as it seems ambiguously intermediate between one and two subjects.

Another advantage of holism about welfare-relevant states is that it helps the state-counting approach respond to Javier-Castellanos's claim that treating pain states as mattering in themselves, rather than as mattering for the subjects who undergo them, amounts to being 'fetishistic' about pain. As he articulates this worry (drawing on arguments in Chappell 2015), "pains are only bad because they are bad for the beings who experience them... [so] placing the emphasis on the pains themselves... gets things exactly backwards." (Javier-Castellanos 2021, p.20). In cases of slight overlap, holists can accommodate the thought that it would be fetishistic to focus on the intrinsic badness of a particular pain by stressing that its badness largely is not intrinsic to it: it matters only insofar as it is incorporated into a complex mind, and renders that mind's total experience unpleasant.²⁶

Is holism about welfare-relevant states, however exactly understood, plausible? For the non-hedonic mental states that might be relevant to welfare (preferences, judgements, intentions, and so on) it is arguably very plausible. Desires, for instance, plausibly would not be the states they are without their connections to other mental states like intentions and beliefs, and there is obviously something holistic about the process of weighing competing preferences, deciding what to do, and acting on that all-things-considered intention.

But the plausibility of holism about hedonic valence - feeling good and feeling bad - is less clear. A defender of holism might point especially at the widespread occurrence of what Bradford (2020, p.239 ff) calls "hurts so good" experiences. These are cases where a usually-unpleasant experience seems to contribute positively to the valence of a total experience: cases like engagement with tragedy, horror, or other sorts of painful art; cases where suffering in

²⁶ It might be objected that pains are still being treated as mattering independently of subjects in cases of complete overlap, like that between you and your head. But this is misguided: subjects with all the same welfare-relevant mental states overlap *as subjects*, sharing their reflections, their agency, their intentions and values or the holistic pattern of their feelings and desires. They share what makes them matter as individuals, and so it is not fetishistic to think of them together for moral purposes: indeed, it might be fetishistic to insist that we count them separately just because they are not technically identical, by strict metaphysical criteria

order to accomplish something makes the struggle feel more alive and the achievement feel more worthwhile; cases where someone seeks out an intensely unpleasant feeling 'just to feel something'.²⁷ As Bloom says, "under the right circumstances and in the right doses, physical pain and emotional pain, difficulty and failure and loss, are exactly what we are looking for." (Bloom 2021; cf. Rozin et al. 2013)²⁸

On the other hand, an opponent of holism might point to the more usual cases where our suffering seems to trace entirely to the intrinsic badness of one intense pain, which feels bad seemingly whatever we think or decide or intend about it. In response, the holist might simply note that in such cases we are unfortunately unable to prevent that one intense pain from causing our total experience to have a negative valence. Nevertheless, the holist might insist, if the overall mental context were sufficiently different (perhaps after intense meditative training, or neurotechnological enhancement) it would be possible to feel *that very pain* and yet be indifferent to it. Both sides can likely offer analyses to explain away the examples offered by the other, so without trying to decide the issue one way or the other, we will stick with the conditional: to the extent that hedonic valence is holistic in its determinants, the tension between subject-counting and state-counting intuitions can be resolved.

If our approach to reconciliation succeeds, it suggests that state-counting approaches are more able than their opposite to accommodate divergent intuitions, under a greater range of metaphysical views. Subject-counting views must either violate some intuitions that

²⁷ For discussions of the 'paradox of painful art' see esp. Feagin 1983, Ridley 2003, Smuts 2009, Bantinaki 2012.
²⁸ Holistic claims about hedonic valence and about preferences could be mutually supporting, insofar as hedonic valence is entangled with or even reducible to conative states. It might just be that how good or bad something makes us feel is often heavily inflected by whether we're getting what we wanted, or having our wishes and efforts frustrated. Or it might be that states only have valence because and insofar as they involve strong desires - that pleasure is good because it's a state we can't help but want to continue, and suffering is bad because it's a state that we can't help but want to end. If so - and if our wantings are inherently tied to a relatively holistic network of psychological attitudes - that would support seeing the value or disvalue of a particular pleasure or displeasure as importantly extrinsic to the experience itself.

state-counters save (about puzzles of multiplication, or constitutive panpsychism) or else be committed to particular metaphysical positions that rule out state-sharing in these cases (e.g. rejecting manyism, rejecting constitutive panpsychism). This is not a decisive objection to subject-counting: many people might accept those constraints on their metaphysics very happily. Indeed, writers like Simon seek to use an implicit subject-counting assumption to motivate those constraints. But the greater flexibility of the state-counting approach can still be regarded as on balance a virtue.

Section 5: Implications for Other Questions

Finally, what about the other questions connected to the Value-Counting question? First, consider the Hedonic Emergence question: must a valenced composite experience be made of valenced parts? Holism about valence suggests, without establishing, a negative answer to this question: a valenced state could be composed entirely out of unvalenced parts. Holism suggests this because it makes the determinants of a total experience's valence something much more structural than just the presence of valence in its parts; it does not, however, strictly establish it, because it remains possible that valence in the parts is a necessary ingredient even if not the sole ingredient. One implication is that holism about valence may serve to moderate the moral implications of panpsychism by undermining any direct inference from humans having valenced complex experiences to their microscopic parts having valenced simple experiences.

Second, consider the Hedonic Alignment question: must the valence of a composite experience match the valences of its parts? Any sort of holism suggests that alignment between the welfare-relevant states of part and whole is likely to be limited, particularly if the degree of overlap between the whole and any particular part is small. Suppose, for example, that the welfare value of the whole's total experience depends on its hedonic valence, which depends on its overall structure, with the valence of any component experience influencing this modestly but not decisively. With large parts (e.g. two conscious half-minds) we can expect some significant degree of alignment (the whole cannot be pleased if both its halves are displeased, etc.) but with small parts, like individual ant-minds in a colony-mind, there need be no significant degree of alignment even in aggregate: if the colony can be pleased or displeased, it might be so even if each individual ant felt the opposite.²⁹ This suggests that we cannot be blithely optimistic about the value alignment of any sort of technological hive mind composed of human minds or their descendents: the whole is not guaranteed to feel the same way as its parts.

Finally, what about person-affecting and impersonal theories of welfare? It is important to note that, while state-counting/impersonal theories and subject-counting/person-affecting theories seem like natural pairs, these pairs can also come apart. For instance, a person-affecting view might endorse the state-counting approach as follows: all value comes from what is better or worse for identifiable people, but when those people share states, the value associated with them combines non-summatively (cf. Sutton 2014). This view would retain the characteristic implications of the person-affecting view for debates in population ethics. Likewise, an impersonal view might endorse the subject-counting view as follows: it is impersonally good for as many subjects as possible to undergo as much pleasure and as little displeasure (or whatever constitutes welfare) as possible, regardless of whether they share that pleasure/displeasure with other subjects. This view would retain the characteristic implications of the impersonal view for debates in population ethics of the impersonal view for debates in population soft the presonal view for debates in population soft the presonal view for debates in population as not pleasure (or whatever constitutes welfare) as possible, regardless of whether they share that pleasure/displeasure with other subjects. This view would retain the characteristic implications of the impersonal view for debates in population ethics. Nevertheless, to the extent that there is an obvious affinity between

²⁹ Ned Block's Nation-Brain thought experiment (1978), where a billion human citizens collectively implement the functional architecture of a single human brain, offers an extreme example of the disconnect between attitudes, at least, at the levels of whole and part. Indeed, the psychological independence of the levels is sufficient that it's not a clear case for state-sharing at all (though see Roelofs 2019, pp.190-198).

these two pairs of views, our argument for the state-counting view may lend some support to an impersonal view over a person-affecting one.

Conclusions

In this paper, we introduced an ethical question that arises if welfare-relevant mental states can be shared between subjects: the Value-Counting question, of whether to make evaluations based on counting mental states or on counting their subjects. We then outlined how this question connects to other questions in ethics and metaphysics, including questions about the nature of welfare and about the composition of consciousness. We also discussed a variety of putative examples that generate opposite intuitive answers to the Value-Counting question, and we canvassed some potential solutions.

As we argued here, the most promising solution to this problem is the state-counting approach plus at least moderate holism about welfare-relevant mental states. On this view, what matters is the value of mental states, but many mental states are good or bad primarily or exclusively because of the contribution that they make to other, more complex mental states that include them. This view denies that we should give extra weight to shared states simply in virtue of their being shared, but it allows that many states can cause more positive or negative experience overall in virtue of their being shared.

Of course, these issues will need more research. And even if all the philosophical questions are resolved, many other empirical and practical questions will remain. For instance, can we produce more happiness, or more value, by creating a web of overlapping minds or a set of separate minds, all else equal? Given the importance these questions may take on as

technology expands the range of potential mind-structures we can create or become, they will

deserve careful attention moving forward.

References

Albantakis, Larissa, Barbosa, Leonardo, Findlay, Graham, Grasso, Matteo, Haun, Andrew, Marshall, William, Mayner, William, Zaeemzadeh, Alireza, Boly, Melanie, Juel, Bjørn, Sasail, Shuntaro, Fujii, Keiko, David, Isaac, Hendren, Jeremiah, Lang, Jonathan, and Tononi, Giulio. (2022). "Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms." *Quantitative Biology: arXiv:2212.14787v1*

Arrhenius, Gustaf. (2003). "The Person-Affecting Restriction, Comparativism, and the Moral Status of Potential People." *Ethical Perspectives* 10 (3-4):185-95.

Bader, Ralf. (2022). "Person-Affecting Utilitarianism." In G. Arrhenius (ed.), *The Oxford Handbook of Population Ethics*, Oxford University Press: 251-270.

Bantinaki, Katarina. 2012. "The Paradox of Horror: Fear as a Positive Emotion." *The Journal of Aesthetics and Art Criticism* 70(4): 383-392.

Blackmon, James. (2016). "Hemispherectomies and Independently Conscious Brain Regions." *Journal of Cognition and Neuroethics* 3 (4): 1–26.

Blackmon, James (2021). "Integrated Information Theory, Intrinsicality, and Overlapping Conscious Systems", *Journal of Consciousness Studies* 28 (11–12): 31–53.

Block, Ned. (1978). "Troubles with Functionalism." In C. W. Savage (ed.), *Perception and Cognition: Issues in the Foundations of Psychology*. Minneapolis: University of Minneapolis Press.

Bloom, Paul. (2021). "The Sweet Spot: The Pleasures of Suffering and the Search for Meaning." Harper Collins.

Bostrom, Nick. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

Bradford, Gwen. (2020). "The badness of pain." Utilitas 32, 236-252.

Briggs, R. A., and Nolan, Daniel. (2015). "Utility Monsters for the Fission Age." *Pacific Philosophical Quarterly* 96 (2): 392-407.

Brüntrup, Goedehard, and Jaskolla, Ludwig, eds. (2017). *Panpsychism: Contemporary Perspectives*. Oxford: Oxford University Press.

Buchanan, Jed, and Roelofs, Luke. (2019). "Panpsychism, Intuitions, and the Great Chain of Being." Philosophical Studies. 176 (11): 2991-3017.

Burke, M. (1994). "Dion and Theon: An Essentialist Solution to an Ancient Puzzle." *Journal of Philosophy* 91 (3): 129–139.

Carls-Diamante, S. (2017). "The Octopus and the Unity of Consciousness." Philosophy of Biology 32 (6): 1269–1287.

Carls-Diamante, S. (2022). "Where Is It Like to Be an Octopus?" *Frontiers in Systems Neuroscience*: <u>https://doi.org/10.3389/fnsys.2022.840022</u>

Chalmers, D. (2015). "Panpsychism and Panprotopsychism." In Y. Nagasawa and T. Alter (eds.), *Consciousness in the Physical World: Essays on Russellian Monism*. Oxford: Oxford University Press.

Chalmers, David. (2019). "Extended Cognition and Extended Consciousness." In Matteo Colombo, Elizabeth Irvine & Mog Stapleton (eds.), *Andy Clark and his Critics*. Wiley-Blackwell: 9-20.

Chappell, Richard Yetter. (2015). "Value receptacles." Noûs 49(2), 322-332.

Clark, Andy. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press.

Clark, Andy, and Chalmers, David. (1998). "The Extended Mind." Analysis 58(1), 7-19.

Cochrane, Tom. (2021). "A Case of Shared Consciousness." Synthese 199 (1-2): 1019-1037.

Dainton, B. (2011). "Review of Consciousness and Its Place in Nature." *Philosophy and Phenomenological Research* 83 (1): 238–261.

Danaher, J., and Nyholm, S. (2021). "Should we use technology to merge minds?" Cambridge Quarterly of Healthcare Ethics 30 (4): 585–603.

Danaher, J., and Petersen, S. (2021). "In Defence of the Hivemind Society." Neuroethics 14 (2): 253–267.

Dominus, Susan 2011. "Could Conjoined Twins Share a Mind?" *New York Times Magazine*, <u>https://www.nytimes.com/2011/05/29/magazine/could-conjoined-twins-share-a-mind.html</u> (retrieved September 2022).

Feagin, Susan. (1983). "The Pleasures of Tragedy." *American Philosophical Quarterly* 20 (1): 95–104.

Friedman, Daniel A., and Eirik Søvik. (2021). "The Ant Colony as a Test for Scientific Theories of Consciousness." Synthese. Synthese (2021) 198:1457–1480

Gauthier, David. (1962). Practical Reasoning. Oxford, Clarendon Press.

Gazzaniga, M., Bogen, J., and Sperry, R. (1962). "Some Functional Effects of Sectioning the Cerebral Commissures in Man." *Proceedings of the National Academy of Sciences* 48 (2): 17–65.

Goff, Philip. (2017). *Consciousness and Fundamental Reality*. New York: Oxford University Press.

Hershenov, D. B. (2013). "Who doesn't have a problem of too many thinkers?" *American Philosophical Quarterly*, 50(2), 203–208.

Hinton, Timothy. (2009). "Rights, Duties and the Separateness of Persons." Philosophical Papers 38 (1): 73-91.

Hirstein, William. (2008). "Mindmelding: Connected Brains and the Problem of Consciousness." *Mens Sana Monographs*, 6 (1): 110–130.

Hirstein, William. (2012). *Mindmelding: Consciousness, Neuroscience, and the Mind's Privacy*. Oxford University Press.

Huebner, Bryce. (2014). *Macrocognition: A Theory of Distributed Minds and Collective Intentionality*. Oxford University Press.

Javier-Castellanos, Arturo. (2021) "Should the Number of Overlapping Experiencers Count?" *Erkenntnis*, <u>https://doi.org/10.1007/s10670-021-00427-4</u>

Johnston, Mark. (2017). "The Personite Problem: Should Practical Reason be Tabled?" *Noûs* 51 (3): 617–644.

Johnston, Mark. (2021). "The Subject and its Apparatus: are they Ontological Trash?" *Philosophical Studies* 178 (8): 2731–2744.

Kang, Shao-Pu. (2022). "Shared Consciousness and asymmetry." Synthese 200: 413.

Kriegel, Uriah. (2017). "Dignity and the phenomenology of recognition-respect." In J. Drummond, & S. Rinof- ner-Kreid (Eds.), *Emotional experiences: Ethical and social significance*, New York: Rowman & Littlefield: 121–136.

Klein, Colin and Barron, Andrew. (2016) "Insects have the capacity for subjective experience." *Animal Sentience* 9 (1).

Klein, S. B. (2015). "The feeling of personal ownership of one's mental states: A conceptual argument and empirical evidence for an essential, but underappreciated, mechanism of mind." *Psychology of Consciousness: Theory Research and Practice*, 2(4), 355–376.

Langland-Hassan, Peter. (2015). "Introspective Misidentification." *Philosophical Studies* 172 (7): 1737–1758.

Lewis, David. (1993). "Many, but Almost One." In *Ontology, Causality, and Mind: Essays in Honour of D.M. Armstrong*, eds J. Bacon, K. Campbell, and L. Reinhardt, Cambridge: Cambridge University Press: 23–45.

Lyreskog, David,Zohny, Hazem, Savulescu, Julian, and Singh, Ilina. (2023). "Merging Minds: The Conceptual and Ethical Impacts of Emerging Technologies for Collective Minds." Neuroethics 16:12 <u>https://doi.org/10.1007/s12152-023-09516-3</u>

Mathews, Freya. 2003. For Love of Matter: A Contemporary Panpsychism. Albany: State University of New York Press.

Miller, Greg. (2017). "Can Subjects Be Proper Parts of Subjects? The De-combination Problem." *Ratio* 30 (2): 1–18.

Mørch, Hedda Hassel. (2019). "Is the Integrated Information Theory of Consciousness Compatible with Russellian Panpsychism?" *Erkenntnis* 84 (5): 1065-1085.

Noonan, Harold. (2010). "The Thinking Animal Problem and Personal Pronoun Revisionism." *Analysis* 70 (1): 93-98.

Nozick, Robert. (1974). Anarchy, State, and Utopia. Oxford, Basil Blackwell.

Oizumi, Masafumi, Albantakis, Larissa, and Tononi, Giulio. (2014). "From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0." PLOS Computational Biology 10 (5). Online.

Olson, Eric. (2003). "An Argument for Animalism." In *Personal Identity*, eds R. Martin and J. Barresi, Malden, MA: Blackwell: 318–35.

Parfit, Derek. (1984). Reasons and Persons. Oxford University Press.

Pinto, Y., D. Neville, M. Otten, P. Corballis, V. Lamme, E. de Haan, N. Foschi, and M. Fabri. (2017). "Split Brain: Divided Perception but Undivided Consciousness." *Brain: A Journal of Neurology* 140 (5): 1231–1237.

Pyke, J. (2014). *Twin life: Sharing mind and body*. Twin Life Productions in association with Canadian Broadcasting Company

Pyke, J. (2017). *Inseparable: Ten years joined at the head*. Curious Features in association with Canadian Broadcasting Company

Rawls, John. (1971). A Theory of Justice. Cambridge, Mass., Harvard University Press.

Ridley, Aaron. (2003). "Tragedy." In J. Levinson (ed.) *The Oxford Handbook of Aesthetics*. New York, NY: Oxford University Press: 408–421.

Roelofs, Luke. (2016). "The Unity of Consciousness, Within and Between Subjects." *Philosophical Studies* 173 (12): 3199–3221.

Roelofs, Luke. (2019). *Combining Minds: How to Think about Composite Subjectivity*. Oxford University Press.

Roelofs, Luke. (2020). "Can We Sum Subjects? Evaluating Panpsychism's Hard Problem." In W. Seager (ed.), *The Routledge Handbook of Panpsychism*. New York: Routledge: 245-258.

Roelofs, Luke. (2022). "No Such Thing as Too Many Minds." *Australasian Journal of Philosophy*.

Rozin, Paul, Guilot, Lily, Fincher, Katrina, Rozin, Alexander, and Tsukayama, Eli. (2013). "Glad to be sad, and other examples of benign masochism." Judgment and Decision Making 8 (4): 439-447.

Schechter, E. (2015). "The Subject in Neuropsychology: Individuating Minds in the Split-Brain Case." *Mind and Language* 30 (5): 501–525.

Schechter, Elizabeth.. (2018). *Self-Consciousness and "Split" Brains: The Minds' I.* Oxford: Oxford University Press.

Schwitzgebel, Eric. (2020). "How Robots and Monsters Might Destroy Human Moral Systems." In *A Theory of Jerks and Other Philosophical Misadventures*, MIT Press: 101-105.

Seager, William. (1995). "Consciousness, Information and Panpsychism." *Journal of Consciousness Studies* 2–3: 272–288.

Sebo, Jeff. (2015-a). "The Just Soul." Journal of Value Inquiry 49 (1-2):131-143.

Sebo, Jeff. (2015-b). "Multiplicity, Self-Narrative, and Akrasia." *Philosophical Psychology* 28 (4):589-605.

Simon, Jonathan. (2017). "The Hard Problem of the Many." *Philosophical Perspectives* 31 (1): 449–68.

Skrbina, David. (2005). Panpsychism in the West. Cambridge, MA: MIT Press.

Slote, M. 1982. "Goods and Lives". Pacific Philosophical Quarterly 63: 311-26.

Smuts, Aaron. 2009. "Art and Negative Affect." Philosophy Compass 4 (1): 39-55.

Sutton, Catherine. (2014). "The Supervenience Solution to the Too-Many-Thinkers Problem." *Philosophical Quarterly* 64 (257): 619–39.

Temkin, Larry (2014). "Rethinking the Good- A Small Taste." *Law, Ethics and Philosophy* 2: 77–78.

Unger, Peter. (1980). "The Problem of the Many." Midwest Studies in Philosophy 5 (1): 411-68.

Unger, Peter. (2004). "Mental Problems of the Many." In *Oxford Studies in Metaphysics*, Volume 1, ed. Dean W. Zimmerman, Oxford: Clarendon Press: 195–222.

Velleman, J. D. 1991. "Well-Being and Time," Pacific Philosophical Quarterly, 72.1: 48-77.

Vetlesen, Arne. 2019. Cosmologies of the Anthropocene: Panpsychism, Animism, and the Limits of Posthumanism. Taylor & Francis.

Vold, Karina. (2015). "The parity argument for extended consciousness." *Journal of Consciousness Studies* 22: 16-33.

Zimmerman, Dean. (2010). "From Property Dualism to Substance Dualism." *Aristotelian Society Supplementary Volume* 84/1: 119–50.